

Net Asset Value Estimation

The invention relates to a method and system for estimating the net asset value (NAV) of a fund.

5

Managers of mutual funds typically publish the NAV (Net Asset Value) of funds not in real time but only on a daily basis. Normally, the NAV is computed at the end of each day on the basis of the closing values of the fund's underlying holdings (securities etc). Investors who may be interested in buying or selling funds therefore have to base their investment decisions on information which is at least one day old and which remains static throughout the day. Moreover, the normal rule is that buy and sell orders are not acted upon immediately but are deferred until the opening of business on the following day. Thus, by the time that the investor's order is actually acted upon, the information on which it is based is at least two working days old. In periods of high volatility within the financial markets, this delay can make it extremely difficult for investors to make effective decisions.

10
15

Investors' problems are compounded in that managers of mutual funds do not normally disclose, on a daily basis, what the underlying securities are that make up the fund. Holdings data may be available on a quarterly or sometimes on a monthly basis, but by the time the information is available it is too out of date to be of much assistance. The lack of information on the underlying holdings make it difficult or impossible to calculate the current or projected future NAV of a fund on the basis of the actual or projected individual values of the underlying holdings. It is of course possible crudely to predict the current or future NAV of a fund purely on the basis of the historical day by day NAV of

20
25EL853250470US

that fund, but in volatile market conditions such estimates are liable to be extremely poor.

5 It is an object of the present invention to provide a system and method for generating an improved estimate, preferably in real time, of the net asset value of a fund.

10 It is a further object to provide a system and method of evaluating the net asset value of a fund which provides a more secure base on which investors can take investment decisions.

According to a first aspect of the present invention there is provided a method of estimating the net asset value of a fund, comprising:

- 15 (a) obtaining:
- (i) historical index values for a plurality of market indexes;
 - (ii) current index values for the said market indexes; and
 - (iii) historical net asset values for the said fund;
- 20 (b) building a model which defines a compound index in terms of the historical index values, the model being characterised by model coefficients;
- (c) optimizing the model by adjusting the coefficient values to fit the compound index to the historical net asset values; and
- (d) estimating the net asset value of the fund by applying the optimized model to the current index values.

25 According to a second aspect of the present invention there is provided a system

for estimating the net asset value of a fund, comprising:

(a) means for obtaining or storing:

- (i) historical index values for a plurality of market indexes;
- (ii) current index values for the said market indexes; and
- (iii) historical net asset values for the said fund;

(b) means for building a model which defines a compound index in terms of the historical index values, the model being characterised by model coefficients;

(c) means for optimizing the model by adjusting the coefficient values to fit the compound index to the historical net asset values; and

(d) means for estimating the net asset value of the fund by applying the optimized model to the current index values.

By making use of this invention, as described, investors are provided with a considerable advantage in terms of time-to-market, in comparison with those who wait for the usual publication of the daily fund valuations.

The invention makes use of data about the past histories of fund NAVs, and also market indexes, as well as real-time values for the same market indexes. The past dependence of each fund on each of the indexes is preferably determined by means of a multiple linear regression (MLR) algorithm, with the estimate being carried out assuming the computed linear dependence to hold also for the future. Using such an algorithm, the system preferably produces, in addition to a real time NAV estimate, an indication of the reliability of the estimate, and a confidence interval for the estimate.

In the preferred embodiment, a set of indexes suitable to be used within an MLR model for each fund is automatically selected from a large number of candidate indexes. This is preferably carried out by applying an MLR analysis on each fund history, with respect to each of the indexes, in order to identify a smaller set of indexes which, combined with weightings computed out of the MLR coefficients, are best able to describe the specific fund historical behaviour.

10 Preferably, the number of indexes accepted for each fund is constrained to a small number, for example ten. The purpose of the selection is to discover a small set of indexes for each fund from which a good MLR model may be built while at the same time (for reasons of efficiency) trying to keep the number of indexes used low.

15 The market indexes used within the present invention may any of the conventional indexes used to track market performance, or some aspect of it. This may include not only conventional stock indexes (such as the FTSE 100 in London) but also other indexes such as currency indexes, currency exchange rates, futures and so on.

20 The invention may be carried into practice in various ways and one specific embodiment will now be described, by way of example, with reference to Figure 1 which shows, in schematic form, the preferred system for estimating the net asset value of a fund.

25 The system illustrated in Figure 1 essentially consists of a front end (a user

application 10) and a back end (the rest of the system). The back end obtains the necessary information from external sources to calculate the current and/or future NAV of one or more funds, with that information then being passed to the user application for display or use in any appropriate way as required by the application. Typically, the system will automatically be estimating the NAV of several different funds at once.

For each fund under consideration, historical NAV information – preferably on a day by day basis – is collected and stored within a NAV history database 12. The information in the database is kept up to date by uploading, daily, the NAV for each fund as published by the respective fund managers. For each fund, the database associates a fund identifier to the sequence of dates and the corresponding fund NAV as at each date. Typically, daily NAV values are stored for at least the past three months.

For ease of subsequent computation, the NAV histories are adjusted for all of the events influencing daily NAV variations, other than the effects of the markets themselves. So, for example, after each fund dividend, the corresponding fund history is adjusted to simulate a re-investment of the dividend amount. More general adjustments or corrections may be made, as required, to ensure that the historical time series for each fund closely corresponds with the actual market performance of that fund.

Historic values of a number of different market indexes (possibly including currency exchange rates, and futures) are stored within an index history database 14. This database associates an index identifier for each of the stored indexes with the corresponding sequence of dates and index values. Typically,

daily index values for at least the last three months are stored. This historical information is preferably updated daily with the close values of each of the indexes as published by the relevant stock markets and/or index issuers institutions.

5

There may be a very large number of indexes stored within the index history database 14, not all of which will necessarily be useful in calculating current values for each of the funds contained within the NAV history database 12. To reduce the amount of computation required when real time NAV estimates are prepared, each fund in the NAV history database 12 is associated with only a sub-set of those funds contained within the index history database 14. The fund/index associations are maintained within an associations database 18, the contents of which are generated by a model builder 20.

10

The role of the model builder 20 is to define a small set of indexes which are suitable to build a good model for each of the funds within the NAV history database 14. In order to select suitable indexes, the model builder makes use of an MLR (multiple linear regression) engine 22; for each fund in the database, an MLR analysis of the NAV history is iteratively carried out, starting from a pre-defined large set of index histories taken from the history database 14.

15

20

In theory, the "large set" of histories could comprise the entire contents of the index history database. In practice, however, for the sake of efficiency a pre-selection is statically performed, including in the starting set the main stock, bond and money market indexes. This initial pre-selection of indexes is stored in the associations database 18, and is manually updated at regular intervals, for example once a quarter.

25

In a first step of the procedure, the MLR engine 22 checks for multiple co-linearity between the indexes and, if this is found, eliminates the offending indexes from the set. Thus, indexes which can be very well approximated by a combination of other indexes will be removed (leaving such indexes in place could affect the soundness of the subsequent computation). The test for multiple co-linearity is performed by using the MLR engine 22 to regress every index value history against all of the others. If the resulting adjusted R^2 value (see below) is greater than a certain figure (for example 0.95), the main computation is not carried out. The model builder reports that fact back to the application 10, and the computation is aborted. Then, the model builder iteratively removes from the set the index with the lowest MLR, and re-invokes the MLR engine for a new analysis based on the reduced set. This stepwise refinement ends when the set is reduced to some required number of indexes, as required by the application 10, or when the quality of the model would otherwise be reduced below some application-defined threshold. The model quality is preferably measured by the quantity adjusted R^2 - see below - which should preferably not fall below 0.8.

The resulting set of indexes is then associated with the corresponding fund within the associations database 18, ready to be used in the NAV estimation procedure, described below.

More formally, the model builder 20 runs a procedure which may be described by the following pseudo code:

```
Model Building Procedure  
(input: SetCard, RLimit)
```

FOR EACH ThisFund IN Funds NAV histories DB

```
5  (
    GET InitialCommonIndexSet from Funds/Indexes DB
    IdxSet = InitialCommonIndexSet
    GET the NAV history of ThisFund from Funds NAV His. DB
    GET the histories of the indexes in IdxSet from Indexes Values His. DB
10 CALL MRL Engine on ThisFund and indexes in IdxSet histories
    NoGoodSet = all the multi-collinearity cases as reported by MLR
    REMOVE all the element of NoGoodSet from IdxSet

15 LOOP UNTIL: Cardinality of IdxSet <= SetCard OR R < RLimit
    (
        CAL MRL Engine on ThisFund and indexes in IdxSet histories
20     GET regression coefficients and adjusted R2 R
        REMOVE from IdxSet the index with the lower coefficient
    )
    IF R < RLimit THEN BREAK
25 ELSE Store IdxSet associated to ThisFund in Funds/Indexes DB
    )
```

30 In the preferred embodiment, the model builder 20 is called by an off-line automatic scheduler (not shown) to rebuild the tables within the associations database 18. If required, the frequency of the rebuild could be defined by the application 10. A typical application 10 might specify monthly rebuilds, each rebuild requiring the calculation of a set of at least ten indexes to create a model having an adjusted R² figure of at least 0.8.

35 When the application 10 requires a NAV estimate for a particular fund, it requests that information from a NAV estimator 24. Alternatively, the NAV estimator may automatically calculate NAV on a periodic basis and may push that information through to the application 10.

The NAV estimator 24 uses the MLR engine 22 to compute the current NAV for a fund, using an estimating algorithm as described in more detail below. The NAV estimator accepts as input the name of the fund, and outputs the current NAV estimate for that fund along with an adjusted R^2 value and a confidence interval. The adjusted R^2 and the confidence interval are passed (together with the estimated NAV) to the application 10, thereby allowing the application to assess the quality of the estimate that has been provided.

The NAV estimator starts by fetching the appropriate set of indexes for the specific fund from the associations database. That information is then used to extract the corresponding index history values from the index history database 14. The index histories, along with the appropriate NAV history from the NAV history database 12, are supplied to the MLR engine 22. The resulting least squares coefficients are combined with the real-time (current) values for the indexes into a NAV estimate for the fund. The current values are obtained, as needed and in real-time, from a commercial feed 16 supplying real time or delayed values for the market indexes of interest. Both subscription-based and on-demand based connection models are acceptable. In the former, the data are pushed from the feed into the system, e.g. Reuters IDN, Bloomberg DDE; in the latter, the data are obtained by the system issuing a query to the feed, e.g. Reuters investors.

The procedure may be defined by the following pseudo code:

25

NAV Estimation Procedure
(input: **TheFund**, Output: NAV-Estimate, R, CI)

GET IdxSet as associated to **TheFund** in Funds/Indexes DB

10

```
IF IdxSet does not exists THEN EXIT
GET the NAV history of ThisFund from Funds NAV His. DB
GET the histories of the indexes in IdxSet from Indexes Value His. DB
CALL MRL Engine on ThisFund and indexes in IdxSet histories

5 GET regression coefficients, adjusted  $R^2$  R and Confidence Interval CI

COMPUTE NAV-Estimate
RETURN (NAV-Estimate, R, CI)
```

10

As explained above, the MLR engine 22 applies multiple regression analysis to the index histories. The MLR engine accepts as input the fund NAV history along with the corresponding associated index histories as created by the model builder and stored within the associations database. The engine then computes the MLR coefficients which, when used in a linear combination of the indexes, produces a compound index which is most similar to the history of the original fund. In other words, the engine computes how much each index contributes, when combined with the others, to explain the behaviour of the fund. The MLR engine also computes two quality indicators: a reliability co-efficient (describing how "good" the model is) and a confidence interval (describing how accurately the combined index is able to reproduce the known fund NAV history).

15

We now turn to a more detailed description of the actual algorithms used by the MLR engine 22. The engine applies Multiple Linear Regression to describe the degree of linear association between a dependent variable Y and a set of p ('almost' independent) variables X_1, \dots, X_p in terms of p multiple correlation coefficients. These coefficients also reflect the relationship between Y and any of the X_i considered on its own.

25

We shall consider the case when Y is the daily quote for a single security and X_i are the daily quotes for p benchmarks along a period of n days.

30

We shall define the quote *data matrix* \underline{X} as:

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{i1} & x_{ij} & x_{ip} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix},$$

5 The vector of dependent variable \underline{Y} as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and an *error vector* \underline{U} as:

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

10 The dependent variable vector \underline{Y} is assumed to satisfy a linear model "error" vector:

$$\underline{Y} = \underline{X} \underline{\beta} + \underline{U}$$

15 where $\underline{\beta}$ is a vector of p unknown correlation coefficients and \underline{U} is the error vector.

A least square estimation can be used to estimate $\underline{\beta}$. One finds that the vector \underline{b} ,

$$\underline{b} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y}$$

20 minimises the "distance" $(\underline{Y} - \underline{X} \underline{\beta})' (\underline{Y} - \underline{X} \underline{\beta})$, thus minimising the error (where the prime symbol ' indicates a transposed matrix).

By using the fitted parameters \underline{b} an estimated expression for \underline{Y} can be built:

12

$$\hat{\underline{Y}} = \underline{X} \underline{b} = \underline{H} \underline{Y}$$

where:

5

$$\underline{H} = \underline{X} (\underline{X}' \underline{X})^{-1} \underline{X}'$$

Given a new set of observations on the X variables $\underline{x}_0 = (x_{01}, x_{02}, \dots, x_{0p})$ the prediction for the dependent variable y_0 is given by:

10

$$y_0 = \underline{b}' \underline{x}_0$$

and the least square residual \underline{e} is defined as

15

$$\underline{e} = \underline{Y} - \hat{\underline{Y}}$$

If the expected value E of the error \underline{U} is described by:

$$E[U_i U_j] = \delta_{ij} \sigma_u^2,$$

20

the variance of \underline{b} may be shown to be:

$$\text{Var}(\underline{b}) = \sigma_u^2 (\underline{X}' \underline{X})^{-1} = \sigma_u^2 \underline{C}^2$$

25

where $\underline{C}^2 = (\underline{X}' \underline{X})^{-1}$.

An (unbiased) estimator of σ_u^2 is:

30

$$s^2 = 1/(n-p) \underline{e}' \underline{e}$$

where 's' is called the standard error of the regression.

Inferences can be made for the individual regression coefficient β_j .

35

A 100(1- α)% confidence interval for β_j is:

$$b_j \pm t_{\alpha/2, (n-p-1)} (s^2 [C^2]_{jj})^{1/2}$$

where $t_{\alpha, \nu}$ is the right tale value of a t-distribution with ν degrees of freedom, corresponding to a confidence level α . It can be shown that the $100(1-\alpha)\%$ confidence interval for the prediction is given by:

5

$$y_0 \pm t_{\alpha/2, (n-p-1)} s (1 + x_0' (X'X)^{-1} x_0)^{1/2}$$

The coefficient of (multiple) determination R^2 is a measurement of the goodness of fit:

10

$$R^2 = SSR / SST$$

where:

$$SST \text{ (total sum of squares)} = \sum_{i=1}^n (y_i - \bar{y})^2 = \underline{Y}' \underline{Y} - n \bar{y}^2$$

15

$$SSR \text{ (regression sum of squares)} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \underline{Y}' \underline{H} \underline{Y} - n \bar{y}^2$$

R^2 represents the proportion of the total variation in \underline{Y} explained by the regression model (it is also known as "explained variance").

20

A better measurement of the goodness of the fit is the *adjusted* R^2 :

$$R'^2 = 1 - \frac{\frac{SSE}{n-p-1}}{\left(\frac{SST}{n-1}\right)} = 1 - \frac{E[\sigma_u^2]}{E[\sigma_y^2]}$$

25

where a penalty function is employed in order to reflect the number p of indexes used in the model. Another indicator of the goodness of fit is the average relative error:

$$\text{Average Relative Error} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{y}_i}{y_i} - 1 \right)^2}$$

5 The algorithm used by the NAV estimator 24 uses Multiple Linear Regression (as described above) to estimate the regression coefficients of each fund with respect to each benchmark (index history). In general, the underlying assumption is that the correlation among time series changes negligibly within the time-scale of the analysis. Typically, since the estimating time frame is one day, this assumption can usually be safely accepted. The number of days of history included within the calculation affects the results. Preferably, a period of ninety days is taken, although it is found in practice that sixty days may be satisfactory. Periods shorter than 30 days should be avoided.

10 MLR applied to a fund NAV history on a set $\{idx_1, idx_2, \dots, idx_p\}$ of index histories, produces a set $\{b_1, b_2, \dots, b_p\}$ of coefficients.

15 Each value NAV_{it} in the fund NAV history can be combined with the corresponding values in the indexes values histories $\{idxV_{1,ji}, idxV_{2,ji}, \dots, idxV_{p,ji}\}$ to produce the best estimation of NAV_{it} .

$$20 \quad NAV_{it} = \sum_{j=1}^p (idxV_{j,ji} * b_j)$$

Thus, on the assumption that the correlation structure of the model (fund NAV vs. indexes values) can be taken as the same for the near future, or at least within the next period (day), an intra-day estimate NAV_{RT} for the fund can be computed using:

25

15 .

$$NAV_{RT} = \sum_{j=1}^P (idxV_{j,RT} * b_j)$$

where $\{idxV_{1,RT}, idxV_{2,RT}, \dots, idxV_{P,RT}\}$ are the actual real-time (or current) values of the indexes in the model.

5

We may study how good the model is by looking at the *adjusted reliability coefficient* (R'^2), also known as the coefficient of multiple determination. This represents the proportion of the overall variation in a time series explained by the indexes, in the time frame taken into account in the MLR. The higher the reliability coefficient (up to 1.0) the higher the predictive power of the benchmarks over the fund, i.e. the more accurate the forecast is. Whenever the MLR description is good enough, it means that we know how a variation in the indexes used reflects into the time series.

15

Whenever a low value (e.g. less than 0.4) of the reliability coefficient is obtained, the benchmark selection should be regarded as useless for forecasting.

20

The forecasted value comes with a confidence interval, representing the value range a defined percentage of forecasting is expected to fall within (this value can be customized).

This is evaluated by computing the population of errors:

$$Error_i = |NAV_{RT,i} - NAV_i|$$

25

That is, the difference between each NAV value of the fund and its estimation computed by combining the regression coefficient and the corresponding index values. The confidence interval is the minimal error value which is greater than given percentage of the population.

30

A reasonable setting for the confidence interval is 80%.

5 In the preferred embodiment, before running the actual model computation
(above) the MLR engine carries out a further check for multiple co-linearity
among the indexes. The test is performed by regressing every index that the
associations database 18 indicates is to be used, against every other. If the
resulting adjusted R^2 value is greater than a certain figure (for example 0.95) the
regression computation is not carried out. The computation is aborted, and the
10 MLR engine sends an error report to the NAV estimator 24, for reporting back
to the application 10.

For completeness, some exemplary implementation details will now be given
for the system illustrated in Figure 1. It will be understood of course that there
15 are a multitude of ways in which the system could be realised in practice, and
the following is not intended by any means to be comprehensive, definitive, or
restrictive of the overall scope of the invention in its broadest sense.

The real-time feed 16 may be implemented by an "off the shelf" product, such
20 as for example the product supplied by Reuters investors. Since the real-time
values of indexes are quite easy to collect, and the number of relevant indexes is
not too high (perhaps a few hundred) multiple feeds and/or products may be
considered. The feed may be interfaced with the system of Figure 1 via any
standard technology such as HTTP queries, DDE links, COM interfaces, API
25 calls, and so on. Data may be provided in any standard format such as ASCII
strings or files, XML, and so on.

The requirements for the databases 12,14 and 18 are quite simple: a single data

structure, a single query schema, and a simple set of update operations. Any relational DBMS would be perfectly suitable. It would also be possible to use a simpler and cheaper technology, such as ASCII or binary flat files.

- 5 The MLR engine is preferably implemented with speed and efficiency in mind. Thus, an efficient programming language such as C, C++ or Delphi is recommended.

10 As there are no tight performance constraints on the model builder 20 or the NAV estimator 24, these may be implemented by means of a high level programming language such as Visual Basic or Java. Both the model builder and the NAV estimator are preferably interfaced with the user application 10, so the implementation of both is preferably encapsulated in a standard component architecture such as Microsoft Com/Com+ or J2EE Enterprise Java Bean.

15 The precise implementation of the user application 10, and the manner in which the NAV estimates are used by that application, is a matter for the application designer. However, a couple of examples may assist.

20 In a simple case, the application 10 comprises a web site having a single function: to allow a user to obtain information on a variety of mutual funds. The web site may include a home page allowing the user to search for and select funds by name, type and/or category. The list of funds meeting the user's criteria is displayed, and by clicking on any fund name the user is taken
25 to a fact sheet page giving further information about the selected fund. Here, among other structural and performance-related information about the fund, the estimated current NAV is presented, as previously described.

This application is implemented within in the Microsoft DNA framework: web pages are ASP running in a Microsoft IIS5 Web Server. The fact sheet page content is dynamically built into an XML data structure by a COM instantiated
5 by the corresponding ASP and implemented in Visual Basic. The COM component invokes the NAV Estimator and mix its output values (NAV estimation and adjusted R^2) in the final XML data structure with other data extracted by a standard SQL database. The ASP then applies the proper XSLT transformation to the XML content, to produce the final HTML page. Adjusted
10 R^2 is presented as a graphical slider on the screen.

In a second alternative implementation, the system is used in conjunction with a portfolio tracking application. The relevant part of the application 10 is a single page, displaying the current content and value of a mixed asset portfolio.
15 This is the type of feature that may be offered by e-brokerage sites to retail customers, or by in-house or external suppliers of solutions to professional portfolio managers or large financial institutions.

When portfolios are constituted only by positions in stocks, the application is
20 able to exploit a real-time data feed for stock prices: individual positions are computed from the stock prices, and are then combined into the total portfolio value. Where there is also a position in funds, such an approach cannot be used, as their value is fixed only once per day. The system of Figure 1 may be used at least partially to fill this gap, with the real time NAV estimation of the
25 funds in the portfolio being used to evaluate the corresponding positions.

Whatever the details of the user application 10, its portfolio computation

component may use the NAV estimates in exactly the same way that they use a real-time data feed for stock prices: the application simply provides a fund identifier and gets back, in return, a current NAV estimated value. The NAV estimate is then combined, in a suitably weighted manner, with the values of the other funds in the portfolio to create a global portfolio estimated value. Standard weighting techniques may be applied according to the respective confidence interval of the NAV estimate and the confidence intervals of the other values with which it is to be combined. A global confidence interval for the overall portfolio estimated value may also be determined.